

Patrick Hanks

Wie man aus Wörtern Bedeutungen macht: Semantische Typen treffen Valenzen¹

Abstract

Wie versteht ein Hörer oder Leser die von einem Sprecher oder Schreiber beabsichtigte Bedeutung? Syntaktische Strukturen sind zu allgemein, um feine Bedeutungsunterscheidungen auszudrücken. Wörter sind oft sehr mehrdeutig, und aufgrund dessen unzuverlässig als „Bedeutungsleitfaden“. Im Gegensatz dazu zeigt die Korpusmusteranalyse, dass die meisten Äußerungen aus Mustern von vergleichsweise geringer Mehrdeutigkeit aufgebaut sind. Daher stellt sich die Frage: Was ist ein Muster? Muster sind häufig verwendete Sprachbausteine, die aus zwei Elementen bestehen: Valenzen und Kollokationen. Während Valenzen relativ stabil sind, sind Kollokationen extrem variabel. In der Korpusmusteranalyse wird eine große Anzahl von Gebrauchsbelegen jedes Wortes studiert, und seine Kollokationen werden, ihren semantischen Typen entsprechend, lexikalischen Sets zugeordnet.

Jedes Wort einer Sprache ist Bestandteil von mindestens einem Muster. Wenn es Teil von mehr als einem Muster ist, können die Bedeutungen seiner Muster meist durch unterschiedliche Kollokations-Präferenzen unterschieden werden.

Kreative Benutzungen sind Abweichungen von normalen Nutzmustern, aber Abweichungen sind selbst regelgeleitet. Daher benötigt man eine Theorie von Normen und Abweichungen. Da die zwei Regelsysteme interagieren, können wir die Theorie als eine „Doppelhelix“ beschreiben.

1. Einführung

Es ist weithin bekannt, dass Textbedeutung kontextabhängig ist. Aber was ist „Kontext“, und woran erkennt man (relevanten) Kontext? In diesem Beitrag untersuche ich den Einfluss von Kollokationen und Valenz auf Bedeutung und Bedeutungsunterscheidungen. Relevante Valenztheorien wurden von Tesnière (1959) und Halliday (1961) formuliert. Für das Englische sind Wortvalenzen weitgehend und intensiv von Herbst et al. (2004), und Kollokationen von Sinclair (Sinclair et al. 1987; Sinclair 1991, 1998), untersucht worden. Im Folgenden werde ich beide Stränge zusammenführen.

Der Ansatz hier ist korpusgesteuert, nicht nur korpusbasiert. „Korpusgesteuert“ heißt, wir betrachten unaufbereitete Korpusdaten, fragen, wie sie auszuwerten sind, und versuchen dann, auf dieser Grundlage theoretische

¹ Diese Arbeit wurde von dem tschechischen Bildungsministerium (MSM 0021620838) und der Grantagentur der tschechischen Republik (P406/2010/0875) als Teil einer korpusgesteuerten Untersuchung von lexikalischen Mustern am Institut für Formale und Angewandte Linguistik, Karls-Universität zu Prag, unterstützt.

Schlüsse zu ziehen. „Korpusbasiert“ bedeutet, dass man von einigen a priori-Annahmen ausgeht und diese dann mit Korpus- oder Internetdaten vergleicht, um stützende Beweise für die Hypothesen zu finden.

2. Valenzen

Oft entstehen durch Valenzen semantische Unterschiede. Das englische Verb *shower* beispielsweise hat zwei Bedeutungen: ‚duschen‘ und ‚überschütten‘. Von den Verben in den folgenden Beispielen ist das erste intransitiv, mit nur einem Argument,² dem Subjekt ((1); A). Das zweite hat drei Argumente, und ist mit zwei verschiedenen Mustern ((2), (3); B, C) assoziiert.

- (1) He showered and dressed quickly.
- (2) He showered her with kisses.
- (3) *He showered kisses on her.*³

Zwischen (2) und (3) gibt es fast keinen Bedeutungsunterschied, sondern höchstens einen unterschiedlichen Fokus. Die unterliegenden Valenzmuster von (1)–(3) sind:

- A. [NP] shower.
- B. [NP₁] shower [NP₂] {with [NP₃]}
- C. [NP₁] shower [NP₃] {on [NP₂]}

Wenn man die Häufigkeiten vergleicht, mit der die jeweiligen Muster auftreten, so fällt auf, dass die große Mehrheit der Korpusbeispiele den Sinn ‚überschütten‘ hat: In einer Stichprobe von 100 Zufallsbeispielen aus dem British National Corpus (BNC) finden sich lediglich 16 mit der Bedeutung ‚duschen‘.

Das Verb *shower* lässt jedoch noch ein weiteres Valenzmuster zu, und zwar D:

- D. [NP₁] shower [NP₂].

Obwohl intuitiv sofort nachvollziehbar, kommt die Variante D sehr selten vor, im BNC (einem Korpus von immerhin einhundert Millionen Wörtern) sogar überhaupt nicht. Käme diese Version tatsächlich vor, würde man (im Englischen) erwarten, dass das direkte Objekt ein Kind oder ein Hund ist, wie in (4).

- (4) *She showers the dog every Sunday.*

² In diesem Beitrag bedeutet ‚Valenz‘ die gesamte Argumentstruktur und ‚Argument‘ ein Valenzelement, d.h. eine Satzrolle wie Subjekt, Objekt, oder Adverbial.

³ Erfundene Beispiele sind, wie hier, kursiv gedruckt, echte (aus einem Korpus entnommene) hingegen nicht. Wenn nicht anders angegeben, ist das Korpus das BNC.

Bedeutsam ist bei diesem Verb, dass durch das Vorhandensein bzw. Nichtvorhandensein eines Adverbials („with [NP₃]“ in (2) oder „on [NP₂]“ in (3)) ein Bedeutungsunterschied erzeugt wird.

3. Einführung: Kollokationen und semantische Typen

Soweit, so gut. Es muss jedoch angemerkt werden, dass sehr viele semantische Unterscheidungen nicht durch Valenzen, sondern durch Kollokationen entstehen. Valenzen allein genügen nicht, Bedeutungen mit Sicherheit auseinander zu halten bzw. zu charakterisieren.

Für jedes Argument eines jeden Musters jedes Wortes gibt es eine ganze Menge von Kollokationen. Die Kollokationen können, ihren semantischen Typen entsprechend, lexikalischen Paradigmen zugeordnet werden. Der theoretische Hintergrund ist die „Generative Lexicon Theory“ (GL; Pustejovsky 1995). Ein kurzer Abriss:

Die GL wurde als ein theoretischer Rahmen zur Kodierung von Wissen über selektionale Präferenzen von Wörtern und ihren Bedeutungen entwickelt. Gemäß dieser Theorie gibt es vier Arten von Ressourcen für jedes Wort:

- i. **Event Structure/Ereignisstruktur:** definiert den „Event Type“ (Ereignistyp) des Satzes oder des Ausdrucks, z.B. Action, Process, State.
- ii. **Argument Structure/Ereignisstruktur:** spezifiziert die Anzahl und die Art der Argumente eines Prädikats.
- iii. **Lexical-Type Structure/Struktur Lexikalischer Typen:** definiert den semantischen Typ eines Wortes in einer hierarchischen Ontologie von Konzepten, z.B. [[Human]], [[Artifact]], [[Vehicle]], [[Concept]]
- iv. **Qualia Structure/Qualia-Struktur:** liefert eine Basis zur strukturellen Differenzierung der prädikativen Kraft eines lexikalischen Elements. Siehe Abschnitt 10 unten.

Sehen wir uns ein Beispiel an. Fragt man, was der „Event Type“ (Ereignistyp) von *shower* (Verb) mit drei Argumenten ist, erhält man die Antwort: „Es kommt darauf an ...“. Typischerweise findet man *Geschenke* ((5), (6), (9)), *Lob* ((7), (10), (11)) oder *Tadel* ((8), (12)) als direktes Objekt oder präpositionales Objekt und einen *Menschen* als drittes Argument (indirektes Objekt).

- (5) Boris showered the woman with presents.
- (6) Rather than the hoped-for cash, they were showered with snuff-boxes and other trinkets, to Leopold's disgust.
- (7) Lauren Bacall, Bianca Jagger ... and Lionel Blair were among the stars who showered him with praise.

- (8) if they ignore the remark or reply negatively they may be accused of rudeness and/or showered with abuse.
- (9) You long to shower gifts on everyone.
- (10) Chinese parents do, of course, shower love and attention on their children.
- (11) European heads of government ... showered telegrams of congratulation on Clinton.
- (12) Despite all the criticisms showered on this model during the past forty years, it still occupies the center of the stage.

Selbstverständlich sind typische Argumente etwas ganz anderes als notwendige Argumente. Eine Korpusanalyse zeigt, dass eine Typifikation sehr oft möglich ist, notwendige Konditionen jedoch selten vorhanden sind. Obwohl Geschenke, Lob und Tadel typisch sind, findet man in diesen Musterargumenten auch Nomen von anderen semantischen Typen, insbesondere dingliche Objekte ((13)–(15)).

- (13) whistling and swearing offends them and they will shower the guilty person with pebbles and gravel until he stops.
- (14) the DC-10 exploded, showering them with debris.
- (15) The eruption showered debris on Pompeii.

In (13) ist das Subjekt menschlich, in (15) ist es ein Ereignis. In (14) ist unklar, ob das Subjekt von „showering“ der Gegenstand („the DC-10“) ist, oder ob es ein Ereignis ist, also die Explosion. Es wäre folgerichtig anzunehmen, dass das Ereignis der „Täter“ ist. Die Beispiele 14 und 15 sind interessant, da in beiden Fällen die Argumente 2 und 3 („them“ – vermutlich Menschen; „debris“ – Fragmente; „Pompeii“ – ein Ort) mit einem Ereignis als Subjekt (Argument 1) korrelieren. Es wäre unwahrscheinlich, wenn ein Ereignis jemanden mit Lob oder Tadel überschüttete. Das Verb *shower* mit unbelebtem Subjekt korreliert normalerweise mit einem dinglichen Objekt. Aber die Verwandtschaft ist nicht symmetrisch, weil ein direktes dingliches Objekt sowohl mit belebten als auch mit unbelebten Subjekten korreliert. Solche Korrelationen unter Argumenten, in Church/Hanks (1990, S. 28–29) als „Triangulation“ bezeichnet, spielen eine große Rolle in der Bedeutungsbestimmung:

Despite the fact that a concordance is indexed by a single word, often lexicographers actually use a second word such as *from* or an equally common semantic concept such as a time adverbial to decide how to categorize concordance lines. In other words, they use two words to *triangulate in* on a word sense. This triangulation approach clusters concordance lines together into word senses based primarily on usage (distributional evidence), as opposed to intuitive notions of meaning. Thus, the question of what is a word sense can be addressed with syntactic methods (symbol pushing), and need not address semantics (interpretation), even though the inventory of tags may appear to have semantic values.

The triangulation approach requires „art.“ How does the lexicographer decide which potential cut points are „interesting“ and which are merely due to chance? The proposed association ratio score provides a practical and objective measure that is often a fairly good approximation to the „art.“ Since the proposed measure is objective, it can be applied in a systematic way over a large body of material, steadily improving consistency and productivity.

Die beschriebenen Zusammenhänge können als Muster mit semantischen Typen wie folgt formalisiert werden:

- E. MUSTER: [[Human₁]]⁴ shower [[Human₂]] with [[Gift, PLURAL]]
 IMPLIKATUR: [[Human₁]] gibt [[Human₂]] viele [[Gift, PLURAL]]
 – Ereignistyp: Geben
 – [[Gift]] heißt allerlei Geschenke
- F. MUSTER: [[Human₁]] shower [[Human₂]] with [[Physical Object, PLURAL]]
 IMPLIKATUR: [[Human₁]] gibt [[Human₂]] viele [[Physical Object, PLURAL]]
 – Ereignistyp: Werfen
- G. MUSTER: [[Human₁]] shower [[Human₂]] with [[Speech Act, MASS]]
 IMPLIKATUR: [[Human₁]] spricht [[Human₂]] viel [[Speech Act, MASS]] aus.
 – Ereignistyp: Sprechen
 – [[Speech Act]] ist oft Lob oder Tadel

Die semantischen Typen repräsentieren „lexical sets“ – paradigmatische Wortgruppen bzw. Gruppen von lexikalischen Elementen –, die eine intrinsische semantische Eigenschaft gemeinsam haben. Im vorliegenden Kontext (als Argument des Verbs *shower*) ist der semantische Typ [[Speech Act]] u.a. den folgenden Substantiven gemeinsam:

praise, accolades, applause, plaudits, abuse, comments, compliments, questions, mail, blessings, telegrams, congratulations, bids, oratorical bouquets, tiny verbal blons.

Paradigmatische Wortgruppen sind nicht begrenzt: Sobald man glaubt, alle Mitglieder einer Gruppe entdeckt zu haben, gibt es immer die Möglichkeit, dass man auf ein weiteres stoßen wird. Ein lexikalisches Element darf aus mehr als einem Wort bestehen, und ein Wort darf in mehr als einer paradigmatischen Gruppe erscheinen.

Semantische Typen werden in einer „flachen Ontologie“ gespeichert (vgl. Pustejovsky/Rumshisky/Hanks 2004). Jeder semantische Typ in der Ontologie wird mit einer Gruppe bzw. einem Set von lexikalischen Elementen ausgestattet sein, basierend auf den im Korpus für jedes Muster gefundenen Beispielen.

⁴ Semantische Typen sind in doppelten eckigen Klammern angegeben.

4. Semantische Typen vs. Semantische Rollen in der Korpusmusteranalyse

Die Korpusmusteranalyse (Corpus Pattern Analysis; CPA) ist eine Methode, Bedeutung auf Wortverwendungen abzubilden. Die Methode ermöglicht die systematische Analyse aller üblichen Gebrauchsmuster einer Sprache. Sie ermöglicht eine Beschreibung aller Wörtmuster, die aus Valenzen und semantischen Typen bestehen, mit Verb oder Adjektiv als Angelpunkt. Bedeutungen sind eher mit Mustern als mit Wörtern assoziiert. Basierend auf dieser Methodologie ist für das Englische bereits „The Pattern Dictionary of English Verbs“ (PDEV; <http://nlp.fi.muni.cz/projects/cpa>) entstanden.

Für die Analyse der 700 bisher vollständig analysierten Verben braucht das Projekt „Corpus Pattern Analysis“ eine Ontologie von nur 200 semantischen Typen. Dies ist teilweise durch eine Unterscheidung zwischen Typ und Rolle möglich. Um diese Unterscheidung zu erklären, betrachten wir das englische Verb *sentence*. Dies hat nur ein Muster (H).

H. MUSTER: [[Human₁]] sentence [[Human₂]] to [[Event = Punishment]]
 IMPLIKATUR: [[Human₁ = Richter]] verurteilt [[Human₂ = Verbrecher]] zu [[Event = {Gefängnisstrafe | Tod}]]
 BEISPIEL: {*Mr Woods*} *sentenced* {*Bailey*} to {*seven years*}.

Nichts in der intrinsischen Semantik der NPs *Mr Woods* und *Bailey* bedeutet ‚Richter‘ oder ‚Verbrecher‘, und nichts in der intrinsischen Semantik der NP *seven years* bedeutet ‚Gefängnisstrafe‘. Diese Ausdeutungen sind semantische Rollen, die vom Kontext zugewiesen werden. Insbesondere die semantischen Rollen der Nomen werden in einem Satz vom Verb zugewiesen.

Semantische Typen sind in doppelten eckigen Klammern geschrieben. Semantische Rollen, wenn vorhanden, sind nach einem Gleichheitszeichen geschrieben.

5. Das Idiomprinzip

Nach Sinclair (1991, 1998) gibt es im Wortgebrauch eine Spannung zwischen einer „terminological tendency“ (terminologischen Tendenz) und einer „phraseological tendency“ (phraseologischen Tendenz). Die terminologische Tendenz ist die Tendenz von Wörtern, in Isolation eine Bedeutung zu tragen. Die phraseologische Tendenz ist die Tendenz von Wortbedeutungen, im Gebrauchskontext aktiviert zu werden. Diese Spannung nennt Sinclair „the idiom principle“ (das Idiomprinzip). Er stellt sich vor, dass zwei grundlegende semantische Tatsachen („two basic facts“) durch die traditionelle Struktur eines Wörterbuchs (Eintrag – Erklärung – Beispiel) verdeckt werden (1998):

- (a) many, if not most, meanings require the presence of more than one word for their normal realization;
- (b) patterns of co-selection among words, which are much stronger than any description has yet allowed for, have a direct connection with meaning.

Diese Spannung ist in vielen alltäglichen Verben und Adjektiven zu beobachten. Im Folgenden werde ich dies am Beispiel des Verbs *admit* nachvollziehen. Terminologisch betrachtet hat dieses Verb nur zwei Hauptbedeutungen: ‚zulassen‘ und ‚eingestehen‘. Zusätzlich gibt es eine Vielzahl verschiedener Implikaturen. Die Hauptbedeutungen von „admit a child into care“ ((16)) und „admit a country to the United Nations“ ((17)) entsprechen einander (‚aufnehmen‘, d.h. eine Art von ‚zulassen‘), aber die Nebenbedeutungen gleichen sich nur teilweise; in einigen Hinsichten sind sie sogar sehr verschieden. In beiden Fällen gibt es Implikaturen von bürokratischen Verfahren, Amtsschimmel und Schreibung, aber in (16) ist impliziert, dass die Kinder krank oder in Gefahr sind, und „care“ heißt in diesem Falle Jugendamtsaufsicht, nicht etwa Sorgsamkeit. In (17) dagegen darf man annehmen, dass die in die Vereinten Nationen eintretenden Baltischen Staaten nach langen Jahren der Unterdrückung „gesund“ und stabil geworden sind.

(16) ... the proportion of older children admitted into care.

(17) The Baltic nations were admitted to the United Nations [in 1991].

Es stellt sich hier die Frage, wie viele lexikografische Unterscheidungen man treffen sollte, und wie fein diese gegliedert sein sollten. Auf diese Frage gibt es keine einzelne einfache Antwort. Die Entscheidung bleibt dem Lexikografen überlassen, der alle Umstände, einschließlich Zweck und Ziel der Arbeit, beachten muss.

Ich zeige hier die prinzipiellen Unterscheidungen in PDEV, mit Häufigkeiten, Implikaturen und Beispielen.

HAUPTBEDEUTUNG: zulassen

A. 10%

MUSTER: [[Human₁ | {Institution = Hospital}]] admit [[Human₂ = Patient]] (to [[Institution = Hospital]] | into {care}) {for [[Activity = {Care | Treatment}]]}

IMPLIKATUR: [[Human₁]] führt bürokratische Maßnahmen zur Zulassung von [[Human₂]] ins [[Institution = Hospital]] aus

BEISPIEL: The hospital admitted 35 people, four of whom had serious injuries.

B. 18%

MUSTER: [[Human 1 | Institution 2]] admit [[Human 2 | Institution 2]]

IMPLIKATUR: [[Human 1 | Institution]] führt bürokratische Maßnahmen zur Zulassung von [[Human 2]] als Mitglied von [[Institution]] aus.

BEISPIEL: The number of working-class boys admitted to grammar schools was disproportionately small | Spain and Portugal were admitted as new members of the DAC.

C. 3%

MUSTER: [{Human = Judge} | {Institution = Court}] admit [[Anything = Evidence]]

IMPLIKATUR: *Law.* [[Human = Judge]] lässt zu, dass [[Anything]] offiziell als Beweismaterial betrachtet wird.

BEISPIEL: Israel's Supreme Court refused to admit evidence allegedly showing that Demjanjuk had been the victim of mistaken identity.

HAUPTBEDEUTUNG: eingestehen

D. 44%

MUSTER: [[Human 1]] admit [NO OBJ] (to [[Human 2]]) {that-CLAUSE | QUOTE = Bad}}

IMPLIKATUR: [[Human 1]] äußert widerstrebend, dass [[Eventuality = Bad]] geschehen ist.

BEISPIEL: General Hakim admitted that the real problem will be fuel.

E. 8%

MUSTER: [[Human]] admit [[Eventuality]]

IMPLIKATUR: [[Human]] äußert widerstrebend, dass [[Eventuality]] wahr ist

BEISPIEL: Some people never do admit defeat.

F. 8%

MUSTER: [[Human]] admit {-ING}

IMPLIKATUR: [[Human]] äußert widerstrebend, dass [[Human]] [[Action = Bad]] begangen hat.

BEISPIEL: She admitted causing the dog unnecessary suffering.

G. 4%

MUSTER: [[Human]] admit [[Action = Bad]]

IMPLIKATUR: [[Human]] äußert widerstrebend, dass er/sie [[Action = Bad]] begangen hat.

BEISPIEL: He admitted eight robberies.

H. 2%

MUSTER: [[Human | Institution]] admit [NO OBJ] {to [-ING]}

IMPLIKATUR: [[Human | Institution]] äußert widerstrebend, dass er, sie oder es [[Action = Bad]] begangen hat.

BEISPIEL: Will people admit to having committed criminal offences, even if they are guaranteed confidentiality?

An den Mustern D–H kann man erkennen, dass ein und dieselbe Implikatur mit mehreren unterschiedlichen Mustern verknüpft sein kann.

Sehen wir uns nun ein Beispiel einer „terminologischen“ Bedeutung an. Auf den ersten Blick erscheint es vielleicht sinnvoll, noch ein weiteres Muster unter der Hauptbedeutung ‚zulassen‘ zu subsumieren, nämlich I.

I. MUSTER: [[Aperture]] admit [[Physical Object | Stuff]]

IMPLIKATUR: [[Aperture]] lässt [[Physical Object | Stuff]] in [[Location]]

Ohne Zweifel ist I idiomatisch wohlgeformt, aber es repräsentiert kein phraseologisches Muster, sondern eine terminologische Bedeutung. Warum? Weil diese Verwendung sowohl selten ist als auch einen hohen Grad semantischer Allgemeinheit aufweist. Obwohl selten, lassen sich doch einige Beispiele mit dieser Struktur finden (z.B. (18)). Solche Beispiele repräsentieren eine terminologische Verwendung des Verbs, nicht eine phraseologische Norm, vgl.:

(18) The skylights in the galleries admit light through angled screens.

6. CPA-Vorgehensweise

Die CPA (Corpus Pattern Analysis; ‚Korpusmusteranalyse‘) ist eine empirische Methode zur semantischen Analyse, die methodische Grundlage von PDEV und ähnlichen Projekten für andere Sprachen. CPA konzentriert sich auf Verben, denn das Verb ist der Angelpunkt des Satzes.

Zur effizienten Korpusmusteranalyse empfiehlt es sich, zuerst in „Sketch Engine“ (Kilgarriff et al. 2004) die statistisch signifikanten Kollokate eines Verbs aufzurufen und zu untersuchen. Wenn ein Kollokat eine bestimmte Bedeutung des Verbs aktiviert, werden alle Korpuszeilen mit dieser Bedeutung einem Muster zugeordnet, d.h. mit der Nummer des entsprechenden Musters getaggt. Aktiviert ein Kollokat mehr als eine Bedeutung, so wird das Verb mit so vielen Musternummern wie nötig getaggt.

Als zweiten Schritt erstellt man zufallsgeneratorisch ein Subkorpus von Beispielzeilen und taggt *jede* Zeile im Subkorpus nach einer der folgenden Kategorien:

- Musternummer⁵ ODER
- Abweichung (Exploitation⁶) von einem vorhandenen Muster ODER
- Auszuschließen (Zitat, Namensteil, Tagging-Fehler, oder einfach unklar).

Wenn man eine Konkordanz das erste Mal öffnet, springen sofort verschiedene Muster ins Auge. Je länger man hinsieht, desto mehr Muster erkennt man. Die Muster basieren auf Kollokationen. Um zu sehen, wie die Kookurenzen von Wörtern Bedeutungen kreieren, müssen wir Kollokationen analysieren. Kollokationen neigen jedoch zur Variabilität, und wenn man versucht, diese Muster zu formalisieren, erkennt man mehr und mehr Ausnahmen. Dabei sind die Grenzen zwischen Mustern und Ausnahmen verwischt. Die meisten Verwendungen eines Wortes können einem Muster zugewiesen werden, aber es gibt fast immer einige (ca. 10%), die ungewöhnlich sind.

7. Was ist ein Muster?

Ein Muster ist eine Festschreibung der Struktur eines Satzes (Valenz) zusammen mit den typischen semantischen Werten (Typen und Rollen) jedes Arguments, realisiert von salienten Kollokaten. Oft aktivieren verschiedene semantische Werte von Argumenten verschiedene Bedeutungen jedes Verbs.

Muster sind kontrastiv, d.h., dass zwei verschiedene Muster eines Verbs in ihrer Argumentstruktur identisch sein und dennoch verschiedene Bedeutungen tragen können. Der Unterschied hängt in solchen Fällen von den semantischen Typen eines oder mehrerer der Argumente ab.

Das Verb *fire*, zum Beispiel, kommt in mehr als 15 Mustern vor. Die fünf häufigsten im PDEV sind die folgenden:

- i. [[Human]] fire [[Firearm]] (at [[Phys Obj = Target]])
- ii. [[Human]] fire [[Projectile]] (from [[Firearm]]) (at [[Phys Obj = Target]])
- iii. [[Human 1 | Institution]] fire [[Human 2]]
- iv. [[Anything]] fire [[Human]] {with enthusiasm}
- v. [[Human]] fire [NO OBJ] (= i oder ii, aber nicht iii oder iv)

⁵ Neue Muster – und ihre Nummern – werden nach Bedarf hinzugefügt.

⁶ Der Begriff „Exploitation“ (Abweichung) wurde zuerst von dem Philosophen H. Paul Grice (1975) benutzt. Er bezeichnet eine Methode der Bedeutungsvermittlung, bei der man darauf verzichtet, explizit zu sagen, was man meint. Ich verwende den Begriff „Abweichung“ im Zusammenhang mit Wortwahl und Musterbedeutung. Sprachbenutzer weichen zu verschiedenen Zwecken von den Regeln normalen Sprachgebrauchs ab, und zwar aus folgenden Gründen:

- a) Um sprachlich schnell und sparsam zu kommunizieren, denn Unterhaltungen werden schnell geführt, und Zuhörer (oder Leser) sind leicht gelangweilt.
- b) Um Neues auszudrücken, z.B. Entdeckungsberichte, neue Hypothesen und dergleichen mehr.

Muster i und ii beschreiben den gleichen Ereignistyp, der Unterschied besteht lediglich darin, dass die Schusswaffe in i bleibt, wo sie ist, während in ii das Projektil in Bewegung gesetzt wird. In FrameNet (<http://framenet.icsi.berkeley.edu>; Baker/Fillmore/Cronin 2003) konstituieren diese zwei Muster ein einziges Frame mit variablen Arten, die verschiedenen Frame-Elemente zu realisieren.

Diese zwei Muster stehen in semantischem Kontrast zu iii, dessen Bedeutung „(aus einem Arbeitsverhältnis) entlassen“ ist. Muster iv stellt einen weiteren Kontrast dar, da das Subjekt nicht auf [[Human]] begrenzt ist, sondern fast alles Mögliche sein kann, z.B. kann eine Einstellung in einen Job jemanden begeistern, aber auch ein philosophischer Gedanke oder ein schönes Lied.

Muster v ist besonders interessant, weil die Bedeutung des Musters von der Abwesenheit eines der Argumente abhängt. Das Nichtvorhandensein des Arguments bezeichnet man als Ellipse. Die Ellipse ist nur unter bestimmten, bislang leider noch unbekanntem Umständen als Alternation zuzulassen. Den Gebrauch eines Verbs mit einem elliptischen oder abwesenden Argument könnte man als ein eigenes Muster bezeichnen. *He fired* kann beispielsweise nur bedeuten: ‚He caused [[Firearm]] to discharge [[Projectile]]‘; und nicht: ‚Er entließ einen/mehrere Arbeitnehmer.‘

8. Die Hypothese von der linguistischen ‚Doppelhelix‘

Nach Hanks (1994) haben Wörter strenggenommen keine Bedeutung, sondern Bedeutungspotenzial. Verschiedene Aspekte dieses Potenzials werden in verschiedenen Kontexten realisiert, entweder durch regelgeleitete Verwendung oder durch Regelverletzung. Die Regelverletzung ist selbst auch regelgeleitet.

Daran kann man erkennen, dass eine Sprache ein System von regelgeleitetem Verhalten ist. In diesem System gibt es nicht ein, sondern zwei miteinander verbundene, interaktive Sets von Regeln:

- a) Regeln, die den Normalgebrauch von Wörtern regulieren;
- b) Regeln, die alle Abweichungen von der Norm (also Regelverletzungen) regulieren.

Beide Arten von Regeln stellen Wahrscheinlichkeiten, nicht Notwendigkeiten, fest.

Jedes Wort wird in einem oder mehreren Gebrauchsmustern (Valenz + Kollokationen) verwendet. Jedes Muster ist mit einer Bedeutung verbunden, wobei eine Bedeutung ein Set von prototypischen Annahmen ist. In der CPA werden Bedeutungen als ‚verankerte Implikaturen‘ ausgedrückt. Die Verankerung wird durch Wiedergabe eines oder mehrerer der seman-

tischen Typen in der Implikatur bewirkt. Um dies zu erklären, betrachten wir noch einmal das Muster und die Implikatur für das Verb *fire*. Wie man erkennen kann, ist die Implikatur eine metasprachliche⁷ Paraphrase der Bedeutung des Musters.

MUSTER: [[Human]] fire [[Firearm]] (at [[Phys Obj = Target]])

IMPLIKATUR: [[Human]] bedingt [[Firearm]], ein [[Projekt]] in Richtung [[Phys Obj = Target]] zu bewegen.

Dieses Muster ist durch Wiedergabe der drei Argumente – [[Human]], [[Firearm]], und [[Phys Obj]] – in seiner Implikatur verankert. Die Argumente erscheinen in der Implikatur und im Muster. Es kann auch Argumente geben (hier: [[Projekt]]), die im Muster nicht ausdrücklich realisiert sind, sondern nur insinuiert. Umgekehrt gibt es auch Muster, die ein Argument enthalten, das in der Implikatur nicht erscheint. Dies ist insbesondere für idiomatische Ausdrücke wie *grasp the nettle* der Fall. Eine *nettle* (Nessel) ist eine Pflanze, aber wenn man von „grasping the nettle“ spricht, redet man nicht von Pflanzen.

MUSTER: [[Human | Institution]] grasp {nettle}⁸

IMPLIKATUR: [[Human | Institution]] fasst einen kühnen Entschluss

Folglich kann man in der CPA einen idiomatischen Ausdruck als ein schlecht verankertes Muster definieren.

Nur wenige Verbmuster sind mit mehr als einer Bedeutung verbunden. Umgekehrt jedoch sind viele Bedeutungen manchmal durch ein ganze Menge von Mustern realisiert, wie schon am Verb *admit* festgestellt werden konnte.

Ein ähnlicher Analyseapparat ist für Adjektive benutzbar. Für prädikative Adjektive ist der Apparat derselbe (mit Einführung einer Kopula) wie für Verben.

MUSTER: [[Human]] be afraid {of [[Anything]]}

IMPLIKATUR: [[Human]] fürchtet [[Anything]]

Für attributive Adjektive muss man Kopfnomen kategorisieren, um verschiedene Bedeutungen zu realisieren:

MUSTER: green [[Phys Obj]]

IMPLIKATUR: [[Phys Obj]] ist grün(farben)

MUSTER: green [[Human]]

IMPLIKATUR: [[Human]] ist unerfahren

Für alle Fälle – Verben wie Adjektive – gilt, dass die Verbindung zwischen Muster und Bedeutung stets prototypisch und statistisch messbar ist.

⁷ Die Metasprache kann jede beliebige Sprache sein (hier: Deutsch).

⁸ Geschweifte Klammern enthalten lexikalische Elemente. Sie sind auch als einfache gruppierende Methode benutzt.

9. Korpusanalyse von Nomen

Der Apparat, den man für die Analyse von Nomen braucht, unterscheidet sich von dem, der für Verben und Adjektive benutzt wird, genauso wie sich ein elektrischer Anschluss von einer Steckdose unterscheidet. Verben sind eng mit Kontext verbunden, aber Nomen sind typischerweise vielmehr auf die Welt bezogen. Daher ist für Nomen die Analyse von Eigenschaften nötig, während Verben Ereignistypen und Argumentstrukturen benötigen.

Many thousands of species of spiders are known (*funnel-web, web-building, orb-weaving, bird-eating, ground-dwelling, giant, huge, large, tiny, poisonous, black widow, camel, redback, trapdoor, wolf, whitetail, crab, tarantula, etc.*).

Some species of spiders *hunt* prey.

Spiders *bite*.

Some species of spiders are *poisonous*.

Many species of spiders *spin* webs, with threads of *strong silk*.

Spiders *lurk* in the centre of their *webs*.

Spiders *control* what is going on in their *webs*.

Spiders have eight *legs*.

Their legs are *thin*, hairy, and long in proportion to body size.

Spiders have *eight eyes*.

Spiders spend a lot of time being *motionless*.

Spiders' *movement* is *sudden*.

Spiders *crawl*.

Spiders *scuttle*.

Spiders are *swift* and *agile*.

Spiders can *run up walls*.

Many people have a dread of (*bate*) spiders.

People *kill* spiders.

In folk taxonomy, *scorpions* and *cockroaches* are often classified together with spiders as creepy-crawly creatures.

Abb. 1: Ein korpusbasiertes kognitives Profil für das englische Nomen *spider*

Nehmen wir das Nomen *spider* (Spinne). Eine statistische Analyse zeigt, dass *scorpion* (Skorpion) und *cockroach* (Kakerlake) zwei der signifikantesten Kollokate von *spider* sind. Als Kollokate stehen sie jedoch in keinerlei besonderer grammatischer Verwandtschaft zu *spider*. Sie finden sich an allen möglichen Stellen in einem Fenster von fünf Wörtern links und rechts von

spider. Solche Korpusbelege gestatten es, für jedes Nomen, das eine Objekt-kategorie denotiert, ein „kognitives Profil“ (oder kognitiven Prototyp; Abb. 1) zu konstruieren. Die Kollokate schließen auch Verben ein, die typischerweise mit *spider* auftreten, sind aber nicht auf diese beschränkt. Das Profil in Abbildung 1 basiert auf Korpusdaten aus dem BNC (British National Corpus) und dem OEC (Oxford English Corpus; 1,5 Milliarden Wörter).

Ein Profil wie dieses hat natürlich nichts mit der wissenschaftlichen Klassifikation von Konzepten zu tun; in gewissem Maße können die zwei Arten der Analyse sogar inkompatibel sein. Andererseits ist ein Profil wie dieses hilfreicher für das Verständnis der Mentalität/Geisteshaltung von Englischsprechern in Bezug auf Spinnen als Informationen, die in traditionellen englischen Wörterbüchern *Spinne* als „arachnid“ definieren.

Ein kognitives Profil wie Abbildung 1 ist mit dem so genannten „lexical conceptual paradigm“ (lcp) kompatibel. Nach Pustejovsky (1995) ist ein lcp

useful for capturing the systematic ambiguities that are so pervasive in language. [...] Nouns such as *newspaper* appear in many semantically distinct contexts, able to function sometimes as an organization, a physical object, or the information contained in the articles within the newspaper.

Für jedes Argument eines Verbs sind nur einige Teile des gesamten lcp irgendeines Nomens relevant. Und in den meisten Fällen wird das Argument eines Verbs von irgendeinem einer großen Anzahl von Nomen realisiert; zusammen konstituieren diese Nomen ein paradigmatisches lexikalisches Set.

10. Qualia

Ein weiterer relevanter Aspekt der Nomenbedeutung wird von Pustejovsky als Qualia (Mehrzahl; Einzahl *quale*) bezeichnet. Dies sind aristotelische Eigenschaften von Konzepten, insbesondere von Nomenkonzepten. Es gibt die folgenden Qualia:

- a) Formales Quale: Was für eine Art von Ding ist es?
- b) Telisches Quale: Wofür ist es?
- c) Konstitutives Quale: Woraus besteht es?
- d) Agentives Quale: Was ist sein Ursprung?

UND (hiermit von mir vorgestellt):

- e) das Axiologische Quale: Ist es gut oder schlecht? (und für wen?)

Das axiologische Quale bedarf der Erklärung. Viele Wörter haben positive/ neutrale/negative Implikationen; *incite*, z.B., hat normalerweise eine negative Konnotation. Obgleich *encourage* ein Fast-Synonym von *incite* ist, stellt sich bei einem Vergleich der beiden Verben heraus, dass *encourage* axiologisch neutral bis positiv ist, während *incite* axiologisch negativ ist.

- You *incite* people to do bad things.
- You can *encourage* someone to write a dictionary (neutral value judgement), but if you say that I *incited* someone to write a dictionary, I'm implying that writing dictionaries is a bad thing to do.

Louw (1993) nennt dieses Phänomen „semantische Prosodie“.

Die semantisch-prosodischen Eigenschaften können etwas kompliziert werden. Ein Beispiel ist das Verb *pander*:

MUSTER: [[Human 1]] *pander* to [[Emotion = Bad | Human 2]]
 IMPLIKATUR: [[Human 1]] gibt [[Emotion = Bad]] von [[Human 2]] nach
 BEISPIELE: *newspapers are **pandering** to people's baser instincts. ... we **pander** to his every whim ... we **pander** to our superiors while neglecting our subordinates.*
 ETYMOLOGIE: von *Pandarus*, Name eines Charakters in Chaucers *Troilus and Criseyde*

Daher sehen wir, dass dieses Verb eine besondere axiologische Komplexität zeigt. Der Gebrauch des Verbs *pander to* (im Vergleich zu *gratify*, *please*, oder *indulge*) impliziert:

- aus Sicht des SPRECHERS: *pandering* ist eine schlechte Handlung.
- aus Sicht des SPRECHERS: die Einstellung oder Emotion des PATIENTEN ist schlecht.
- aus Sicht des AGENS (the *panderer*): *pandering* kann gute Konsequenzen haben, obwohl der AGENS glaubt, dass die Einstellung/Emotion des PATIENTEN schlecht ist.
- PATIENT (the *pandered-to*) hat keine Meinung diesbezüglich.

Das präpositionale Objekt (nach *to*) kann zwei semantische Typen haben: [[Human]] oder [[Emotion]]. Beide sind jedoch immer impliziert; Selektion ist eine Sache des Fokus. Eine solche Alternation kommt in der Argumentstruktur sehr häufig vor.

Auch die Etymologie ist in diesem Falle relevant, weil jemand, der „*pandert*“, sich wie der Charakter bei Chaucer verhält und nicht wie ein griechischer Held.

11. Abweichungen

Nicht alle Korpuszeilen können einem Muster zugeordnet werden. Eine Sprache bietet ihren Nutzern immer die Gelegenheit, kreativ zu sein. Betrachten wir zuerst das Idiom *to clutch at straws*. Sein Ursprung ist das Sprichwort, *a drowning man will clutch at a straw*. Es bedeutet, dass jemand verzweifelt nach Lösungen für ein Problem sucht, obwohl diese höchstwahrscheinlich nicht effektiv sein werden.

- (19) Ms Magee said: „Keith is clutching at straws in an effort to win the argument“.
- (20) Lebanese members of parliament **clutched at** the threads of a peace plan yesterday after a brief period when agreement to end 14 years of civil war seemed within their grasp.

(19) repräsentiert einen normalen Gebrauch des Idioms in seiner kanonischen Form. In (20) wird von dieser kanonischen Form abgewichen – es findet sich in diesem Beispiel keine explizite Realisierung des Wortes *straw*, dennoch ist das Idiom im Hintergrund versteckt; dafür sprechen das kognitive *at* und die Tatsache, dass die libanesischen Parlamentarier nicht im wörtlichen Sinne versuchen, ein dingliches Objekt zu ergreifen. Sie werden als ertrinkende Männer dargestellt, und ‚the threads of a peace plan‘ sind ihre unwirksamen Strohhalme. Die Resonanz des Idioms wird gründlich genutzt,⁹ obgleich eines der wichtigsten lexikalischen Elemente nicht genannt wird. Die Bedeutung des Satzes wird dennoch klar genug, sogar für jemanden, der das Idiom nicht kennt.

Eine Abweichung ist eine absichtliche Abänderung eines etablierten Musters. Sie wird verwendet, um entweder über neue oder ungewöhnliche Dinge/Ereignisse zu sprechen, oder um Bekanntes auf neue, interessante oder unkonventionelle Art auszudrücken. Abweichungen sind Teil der natürlichen Tendenz von Menschen, mit Sprache zu spielen, aber sie sind auch in anderen Situationen zweckdienlich, insbesondere in solchen, in denen es keine etablierte Art und Weise gibt, über eine bestimmte Situation, ein bestimmtes Ereignis oder einen bestimmten Gegenstand zu sprechen.

Man kann im Hinblick auf Abweichungen mindestens vier Punkte anmerken:

- i. Abweichungen sind selten im Vergleich zu Mustern. Der Gedanke an einen Politiker, der nach dem sprichwörtlichen Strohalm greift, ist viel einprägsamer als der Gedanke an einen Politiker, der eine Rede hält.
- ii. Abweichungen sind kognitiv salient. Soziale (oder statistische) Salienz lässt sich als häufige Benutzung definieren (oder ist als solche erkennbar). Ist ein Ausdruck kognitiv salient, so lässt er sich leicht ins Bewusstsein/Gedächtnis zurückrufen.
- iii. Aus diesem Grund sind Abweichungen rhetorisch effektiv. Wie Quintilian und andere Rhetoriker der klassischen Antike vor langer Zeit feststellten, sind Metaphern und andere Figuren (von denen viele phanta-

⁹ Resonanz gilt in der Metaphertheorie als eines der Kriterien für Metaphorizität, vgl. Hanks (2010, S. 140–141): „[...] if one sense of an expression resonates semantically with another sense, it is metaphorical, and if there is no such resonance, it is literal“ sowie Hanks (2006, S. 20): „[...] secondary senses that are classed as metaphors ‚resonate‘ (Black 1962) with some other term (the primary subject) in the immediate context in a text“.

sievolle Abweichungen von Normen sind) viel einprägsamer als der alltägliche Sprachgebrauch, daher ermutigten sie ihre Schüler, diese Figuren zu nutzen.

- iv. Sie machen den Terminus „selektionale ‚Restriktionen‘“ unsinnig. So etwas wie selektionale Restriktionen gibt es nicht; es gibt nur selektionale Präferenzen. Abweichungen sind völlig wohlgeformte, bedeutungsvolle und beabsichtigte Benutzungen von Sprache, obwohl sie außerhalb der selektionalen Präferenzen eines Wortes liegen. Sie werden nicht durch Restriktionen unmöglich gemacht. Ein Problem der Wortanalyse ist die unscharfe Abgrenzung zwischen Normen und Abweichungen; einige Wortverwendungen sind üblicher – normaler – als andere.

Das Bild wird verkompliziert durch verschiedene andere Faktoren, insbesondere:

- v. Manchmal wird eine gelungene Abweichung von anderen Sprachbenutzern aufgegriffen und etabliert sich als Norm. Eine Abweichung von heute könnte eine Norm von morgen sein.
- vi. Domänenspezifische Normen (z.B. der Jargon eines bestimmten Sportes oder von Sportjournalisten) finden sich in allgemeinen Korpora und sind selbst exploitierbar, sobald sie sich etabliert haben. Ausdrücke wie *climb above the defenders* und *fire the ball into the back of the net* sind nicht nur in Fußballjournalismus und Sportreportagen großer Zeitungen zu finden; sie könnten aufgegriffen und in Kontexten außerhalb des Sports metaphorisch genutzt werden.
- vii. Es gibt keine klare Grenze zwischen Normen und Abweichungen. Manche Äußerungen sind extrem normal, andere sind Abweichungs-extreme. Es gibt eine große Grauzone, in der einige Äußerungen weniger normal sind als andere. Eine Abweichung von einer Norm kann ein Einzelfall sein, oder sie kann sich als text- oder domänenspezifische Abweichung etablieren.

Das Ausmaß, in dem eine lexikalische Norm offen ist für Abweichungen, variiert von semantischem Typ zu semantischem Typ und von Wort zu Wort. Das Phänomen der Abweichung ist ein zentraler Aspekt der kreativen, dynamischen Natur von Sprache, so dass es nicht überrascht, dass es sich schwer festmachen lässt. Das Nichterkennen des Phänomens aber hat in der Vergangenheit dazu geführt, dass der Untersuchungsgegenstand Sprache noch komplizierter erschien als er ohnehin schon ist.

Eine Sprache besteht aus einem System von sich stets bewegenden und entwickelnden Regeln, die das linguistische Verhalten steuern: normaler Sprachgebrauch auf der einen Seite, Abweichungen von normalem Sprachgebrauch auf der anderen.

Ein Hörer oder Leser kann zwischen einer echten neuen Abweichung und einer seltenen Norm letztendlich nicht unterscheiden. Ein Ausdruck, der dem Hörer unerhört neu und anschaulich vorkommt, kann in diesem Moment das erste Mal geprägt worden sein. Andererseits könnte der Sprecher diesen Ausdruck schon lange verwendet haben, er könnte ihn vor kurzem aufgeschnappt haben, oder der Ausdruck könnte Teil eines konventionellen Vokabulars einer (dem Hörer) unvertrauten Domäne sein. Dies ändert jedoch nichts daran, dass von den Normen einer Sprache dynamisch abgewichen werden kann und abgewichen wird. Es wäre vielleicht sinnvoll, von Sprecher-/Schreibernormen und Zuhörer-/Leserabweichungen zu sprechen.

12. Verschwommenheit

Die Grenzen aller linguistischen und lexikalischen Kategorien sind verschwommen. Es gibt daher zahllose schwierige Grenzfälle. Anstatt über Grenzen zu streiten, sollten wir jedoch Prototypen identifizieren. Anschließend wird man in der Lage sein, zu entscheiden, was mit was korreliert. Viele Entscheidungen werden offensichtlich sein, während man einige Entscheidungen – insbesondere bezüglich der Grenzfälle – willkürlich treffen müssen wird.

Die Korpuslinguistik hat gezeigt, dass natürliches linguistisches Verhalten stark geregelt ist – und trotzdem verschwommen. Viele Muster sind entdeckt worden; andere werden ohne Zweifel noch entdeckt werden. Alle bisherigen theoretischen Spekulationen müssen im Licht der neuen Korpusdaten neu bewertet werden. Gibt es dennoch einen Daseinszweck für spekulative Linguistik? Meine Antwort: Ja, sofern man sich auf die Interpretation von Daten beschränkt.

- Kontrastive negative Beispiele („Man darf nicht X sagen“) können durch korpuslinguistische Methoden nicht entdeckt werden.
- Sprache ist ein analoges System – oder eine verwirrende Mischung von Analogie und Prädikatenlogik.
- Außerhalb mathematischer Logik gibt es keine scharfen Grenzen.
- Kollokationen sind analoge Sets.
- Kollokationen können – und müssen – gemessen werden, wenn wir die Grundlagen von Bedeutung verstehen wollen.
- Bedeutungen sind holistisch – der Feind ist geistlose Kompositionalität (bzw. grammatikalischer Reduktionismus).

Die meisten lexikalischen Sets sind nicht stabil. Hanks/Jezek (2008) und Jezek/Hanks (2010) bezeichnen sie als schimmernde lexikalische Sets. Ein Beispiel:

MUSTER: [[Human]] attend [[Event]]

In diesem Beispiel ist [[Event]] ein schimmerndes lexikalisches Set. Das „Event“ bzw. Ereignis könnte z.B. *meeting* (Treffen), *wedding* (Hochzeit), *funeral* (Beerdigung) usw. sein. Aber nicht alle Ereignisse gehören zu diesem Set in diesem Kontext. Ereignisse, die nicht zum lexikalischen Set des Verbs *attend* gehören, sind z.B. *thunderstorm* (Gewitter) und *suicide* (Selbstmord) – um nur zwei zu nennen. Auf der anderen Seite passen nicht nur Ereignisse in dieses Paradigma, sondern auch Orte, wo relevante Ereignisse stattfinden; vgl. *attend school*, *attend a clinic*. *School* und *clinic* bezeichnen Institutionen oder Orte, nicht Ereignisse. Allerdings: Man „attends“ a school or clinic, um an den Ereignissen, die dort stattfinden, teilzunehmen.

13. Schlussbemerkung

Die Korpusmusteranalyse (CPA) ist eine Forschungsmethode zur Untersuchung von Valenzen und Kollokationen. Ein Ziel des CPA-Projektes ist es, die Analyse von Valenzen und semantischen Typen von Verbarargumenten zusammenzubringen, um eine Ressource zu schaffen, mit Hilfe derer man Textbedeutung auf Wortverwendung abbilden kann. Diese Ressource besteht aus einem „Muster-Wörterbuch“, zunächst von englischen Verben. Derzeit wird an entsprechenden Projekten für italienische (Pavia) sowie spanische Verben (Universität Pompeu Fabra, Barcelona) gearbeitet.

Ein weiteres Ziel der Korpusmusteranalyse ist es, Bedeutungstheorie empirisch besser zu begründen. Korpusbelege zeigen, dass die meisten Benutzungen eines Wortes sich an ein Muster anpassen. Die Sprachregularitäten sind viel regelmäßiger als man vor Anfang der Korpusanalyse vorhergesehen hatte. Einige Wortverwendungen sind jedoch sehr unregelmäßig – viel unregelmäßiger als man vor Anfang der Korpusanalyse vorhergesehen hatte: Es handelt sich um Abweichungen.

Bedeutungen sind eher mit Mustern assoziiert als mit einzelnen Wörtern. Ein „Muster-Wörterbuch“ beschreibt das Set von Mustern für jedes Verb, und jedes Muster ist mit einer Implikatur verbunden, die die Bedeutung des Musters als ganzes wiedergibt. Die folgenden Elemente konstituieren ein Muster:

- das Verb selbst;
- die Argumente, mit denen es normalerweise auftritt.

Jedes Argument wird entweder als ein lexikalisches Element realisiert oder als ein Paradigma von lexikalischen Elementen. Solche Paradigmen haben gewöhnlich eine semantische Eigenschaft gemeinsam, die als semantischer Typ dargestellt werden kann. In Bezug auf die Qualia-Struktur handelt es sich meist um das „Formale Quale“.

Semantische Typen werden in einer hierarchischen Ontologie gespeichert. Für die Analyse von 700 englischen Verben hat sich eine flache Ontologie von 200 Elementen als ausreichend herausgestellt. Diese kleine Anzahl ist aus zwei Gründen möglich:

- 1) Es ist möglich, die lexikalischen Elemente aufzulisten, ohne semantische Typen zuzuordnen, und diese Möglichkeit wird oft genutzt.
- 2) Semantische Eigenschaften von Nomen, die ihm von dem Verb zugewiesen werden, mit dem es kookkuriert, werden als semantische Rollen ausgedrückt (z.B. „Judge“).

Es gibt keine Einschränkung bezüglich dessen, was als semantische Rolle benutzt werden kann. Der Unterschied ist, dass semantische Typen intrinsische Eigenschaften eines Nomens ausdrücken, während semantische Rollen vom Kontext zugewiesen werden.

Man muss unterscheiden zwischen kanonischen Elementen in einem lexikalischen Set und ad hoc-Elementen, wie z.B. metaphorischen Wortverwendungen. Die gesamte Vorgehensweise ist analog, statistisch und probabilistisch, daher ist es keine Überraschung, dass es unerwartete und seltene Mitglieder von semantischen Typen gibt.

Die ersten Resultate deuten darauf hin, dass die Muster einen großen Einfluss auf die Bestimmung von Wortbedeutung im Text haben. Die Ergebnisse wurden von Cinkova et al. (2010) auf der Tagung „Text, Speech and Dialogue 2010“ in Brno präsentiert.

Literatur

- Baker, Collin F./Fillmore, Charles J./Cronin, Beau (2003): The structure of the Frame-net database. In: *International Journal of Lexicography* 16, 3, S. 281–296.
- Black, Max (1962): *Models and metaphors: studies in language and philosophy*. Ithaca.
- Church, Kenneth W./Hanks, Patrick (1990): Word association norms, mutual information, and lexicography. In: *Computational Linguistics* 16, 1. S. 22–29. [Wieder erschienen in Fontenelle (Hg.) (2008), S. 285–295 und Hanks (Hg.) (2008)].
- Cinkova, Silvie et al. (2010): Can corpus pattern analysis be used in NLP? In: Sojka, Petr et al. (Hg.): *Text, speech and dialogue. 13th International Conference, TSD 2010, Brno, Czech Republic, September 6-10, 2010. Proceedings.* (= *Lecture Notes in Computer Science/Subseries Lecture Notes in Artificial Science* 6231). Berlin u.a., S. 67–74.
- Fontenelle, Thierry (Hg.) (2008): *Practical lexicography. A reader*. Oxford/New York.
- Grice, H. Paul (1975): *Logic and conversation*. In: Cole, Peter/Morgan, Jerry L. (Hg.): *Syntax and semantics, Bd. 3: Speech acts*. New York u.a., S. 41–58.
- Halliday, Michael (1961): Categories of the theory of grammar. In: *Word* 17, 3, S. 241–292.

- Hanks, Patrick (1994): Linguistic norms and pragmatic exploitations, or why lexicographers need prototype theory and vice versa. In: Kiefer, Ferenc/Kiss, Gábor/Pajzs, Júlia (Hg.): *Papers in computational lexicography. Proceedings of the 3rd International Conference on Computational Lexicography (COMPLEX '94)*. Budapest, S. 89–113. [Wieder erschienen in Hanks (Hg.) (2008)].
- Hanks, Patrick (2006): Metaphoricity is gradable. In: Stefanowitsch, Anatol/Gries, Stefan Th. (Hg.): *Corpus-based approaches to metaphor and metonymy*. (= *Trends in Linguistics: Studies and Monographs* 171). Berlin/New York, S. 17–35.
- Hanks, Patrick (Hg.) (2008): *Lexicology: critical concepts in linguistics*. 6 Bde. London u.a.
- Hanks, Patrick (2010): Nine issues in metaphor theory and analysis. In: *International Journal of Corpus Linguistics* 15, 1, S. 133–150.
- Hanks, Patrick/Jezek, Elisabetta (2008): Shimmering lexical sets. In: Bernal, Elisenda/De Cesaris, Janet (Hg.): *Proceedings of the XIII EURALEX International Congress*. Barcelona, S. 391–402.
- Herbst, Thomas et al. (2004): *A Valency Dictionary of English: a corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives*. (= *Topics in English linguistics* 40). Berlin/New York.
- Jezek, Elisabetta/Hanks, Patrick (2010): What lexical sets tell us about conceptual categories. In: *Lexis – E-Journal in English Lexicography* 4 (*Corpus Linguistics and the Lexicon*), S. 7–22. Internet: http://screcherche.univ-lyon3.fr/lexis/IMG/pdf/Lexis_4_Jezek_Hanks.pdf (Stand: 10.12.2010).
- Kilgarriff, Adam et al. (2004): The Sketch Engine. In: Williams, Geoffrey/Vessier, Sandra (Hg.): *Proceedings of the 11th Euralex International Congress*, Lorient, France, July 6–10, 2004. Bd. 1. Lorient, S. 105–115.
- Louw, Bill (1993): Irony in the text or insincerity in the writer? – The diagnostic potential of semantic prosodies. In: Baker, Mona/Francis, Gill/Tognini-Bonelli, Elena (Hg.): *Text and technology: in honour of John Sinclair*. Philadelphia u.a., S. 157–176.
- Pustejovsky, James (1995): *The generative lexicon*. Cambridge, MA u.a.
- Pustejovsky, James/Rumshisky, Anna/Hanks, Patrick (2004): Automated induction of sense in context. In: *Proceedings of the 20th International Conference on Computational Linguistics*. Geneva, S. 55–58.
- Sinclair, John (1991): *Corpus, concordance, collocation*. Oxford u.a.
- Sinclair, John (1998): The lexical item. In: Weigand, Edda (Hg.): *Contrastive lexical semantics*. (= *Amsterdam Studies in the Theory and History of Linguistic Science: Series 4, Current Issues in Linguistic Theory* 171). Amsterdam/New York, S. 1–24. [Wieder erschienen in Hanks (Hg.) (2008)].
- Sinclair, John et al. (1987): *Collins COBUILD English Language Dictionary*. London u.a.
- Tesnière, Lucien (1959): *Éléments de syntaxe structurale*. Paris.